

Library of Congress Pinyin Conversion Project

Conversion and Cleanup Tasks: Status Report December 30, 2003

COMPLETED CLEANUP TASKS

1. Review of files of converted authority records

While converting authority records, OCLC created a number of files containing certain kinds of records for review. For more than five months in late 2000 and early 2001, LC catalogers painstakingly reviewed the thousands of converted authority records in these files and made corrections where warranted.

2. Authority records for undifferentiated personal names

The conversion program did not convert authority records that were coded as undifferentiated personal names. With the able and generous assistance of a dozen cooperating libraries, more than 8400 of these records were evaluated and converted manually. In the process, several thousand new unique and non-unique authority records were created.

3. Double conversion

These two headings were checked to make sure that they did not “double-convert”:

P‘i-hsien (Kiangsu Province, China) converted to Pi Xian (Jiangsu Sheng, China)

T‘eng-hsien (Shantung Province, China) converted to Teng Xian (Shandong Sheng, China)

4. Subject headings and subject subdivisions for regions in China

Some of the subject headings for regions in China converted correctly, but others did not. Therefore, all headings on bib records for regions in China were located, evaluated, and corrected when necessary.

EXAMPLES:

651 -0 \$a Canton Region (China)... [changed manually to Guangzhou Region (China)]

651 -0 \$a Taiyuan Shi Region (China) [changed manually to Taiyuan Region (Shanxi Sheng, China)]

650 -0 ... \$z Sinkiang Uighur Autonomous Region [changed manually to Xinjiang Uygur Zizhiqu]

650 -0 ... \$z Tangshan (Hebei Sheng) Region [changed manually to Tangshan Region (Hebei Sheng)]

650 -0 ... \$z Luoyang (Henan Sheng) Region [changed manually to Luoyang Region (Henan Sheng)]

5. Multi-syllable terms for Chinese jurisdictions

Ten multi-syllable terms for Chinese jurisdictions were to have been joined together by the conversion program when they were identified as being part of a proper name. Some, however, were joined together in other situations. Also, some of the correctly converted terms had to be changed. (For example, T'ai-wan ti ch'ü converted to Taiwan Diqu; this string had to be changed to Taiwan di qu, because the term di qu (地区) in this instance refers to the Taiwan region in general, and not specifically named location.) We scrutinized each bib record on which these ten terms appeared; many records were corrected.

<u>Term</u>	<u>Hits</u>	<u>Needed Correction</u>
diqu	1670	ca. 1100
tequ	88	ca. 55
xingzhengqu	75	ca. 40
zhuanqu	11	2
dujiaqu	1	0
ziran	33	0
zizhiqi	1300	0
zizhiqu	11	0
zizhixian	253	0
zizhizhou	356	0

6. Bogus multi-syllable terms

On Chinese bib records converted in RLIN, the conversion program incorrectly created several multi-syllable generic terms. These are the terms that have been identified and corrected:

<u>Wade-Giles syllables</u>	<u>Converted to</u>	<u>Should be</u>
ti ch'üan	diquan	di quan
ti ch'üeh	diqueh	di que
tu chia ch'ü	dujiaqu	du jia qu
min tsu	minzu	min zu
te ch'üan	tequan	te quan
hsing cheng ch'üan	xingzhengquan	xing zheng quan
tzu chih ch'üan	zizhiquan	zi zhi quan
chuan ch'üan	zhuanquan	zhuan quan

7. Guangzhouese

On Chinese bib records converted in RLIN, the word Cantonese was converted to Guangzhouese when it appeared in subject headings. This term has been manually corrected on all LC records.

EXAMPLES:

650 -0 \$a Guangzhouese dialects [changed manually to Cantonese dialects]

650 -0 \$a Cookery, Chinese \$x Guangzhouese style [changed manually to Cookery, Chinese \$x Cantonese style]

8. Chinese monograph records that were marked for review in the 987 field

Records on which access points required change have been converted or corrected. The term [access not affected] has been added in the 987 \$f subfield of the remaining bib records; those records have been set aside.

9. Chinese serial records that were marked for review in the 987 field

The 900 remaining serial records that were marked for review have been converted.

10. Unconverted IBC serial records

The 516 brief Chinese acquisition records in the LC database have been reviewed and converted.

11. Personal names with religious titles

To the extent possible, authority records and headings for personal names that included religious titles (such as fa shi 法师, da shi 大师, chan shi 禅师) have been identified and converted.

12. Subject headings that were not converted by machine

Working from the four lists of Chinese subject headings that appear on the pinyin home page, CPSO converted subject headings that were not converted by machine, on all but premarc bib records.

13. Syllable sweep for bib records for instrumental music

Unique Wade-Giles syllables were searched in music records in the LC database. All records that appeared to include romanized Chinese were printed out, reviewed, and converted where appropriate. In all, about 1500 records were converted.

14. Syllable sweep for bib records for motion pictures

Unique Wade-Giles syllables were searched in motion picture records in the LC database. All records that appeared to include romanized Chinese were printed out, reviewed, and converted where appropriate. 3000 records were reviewed, and 90 were converted. Since much of the data on these records that appears to be romanized has, in fact, been transcribed from copyright applications, titles proper on motion picture records were almost never converted. In most instances, the pinyin form of a romanized title was given in a 246 field.

15. Names of geographical features (rivers, mountains, deserts, etc.):

The conversion program connected certain generic terms for geographic features (primarily the terms for rivers) for geographic features to the names that preceded them. These generic terms will be identified and separated on authority and bib records, to conform to the romanization guidelines.

EXAMPLES:

pre-conversion WG form machine converted to: change to:

Chang-chiang	Changjiang	Chang Jiang
Huang-ho	Huanghe	Huang He
Chu-chiang	Zhujiang	Zhu Jiang

At the same time, some 20 multi-syllable generic terms which are used in proper names were not connected by the conversion program. They will be identified and joined together when appropriate.

EXAMPLES:

<u>pre-conversion WG form</u>	<u>machine converted to:</u>	<u>change to:</u>
Huang-t'u kao yüan	Huangtu gao yuan	Huangtu Gaoyuan
Ch'ing Tsang kao yüan	Qing Zang gao yuan	Qing Zang Gaoyuan
San-chiang p'ing yüan	Sanjiang ping yuan	Sanjiang Pingyuan
T'a-k'e-la-ma-kan sha mo	Takelamagan sha mo	Takelamagan Shamo
Ch'ai-ta-mu pen di	Chaidamu pen di	Chaidamu Pendi
Su-i-shih yün ho	Suyishi yun he	Suyishi Yunhe
Pa-na-ma yün ho	Banama yun he	Banama Yunhe

16. "Most frequently used" headings

Lily Kecskes systematically converted the [164] "most frequently used" headings on more than 17,000 bib records in the LC database that were not converted by machine. The task took more than 18 months to accomplish. Lily converted each heading on bib records one by one. In the course of her work, she added more than 50 headings to the list that was originally supplied by OCLC at the beginning of the conversion project. A list of the most-used headings that have been converted may be found at the end of this document.

THE REMINING CLEANUP TASKS

1. Headings for Chinese jurisdictions; conventional place names

Almost all authority records and headings for Chinese jurisdictions on Chinese bib records were correctly converted by the machine program. Most of the headings on Korean and Japanese records on RLIN have also been converted. Headings for conventional headings for provinces are now being corrected on non-Chinese and PREMARC records. However, because of the many recent changes to the names and boundaries of Chinese cities and counties, a comprehensive review of these headings is being conducted, and headings for Chinese jurisdictions are constantly being updated.

This is complex and time-consuming work that will doubtless take months, or perhaps years, to accomplish. When the name of one jurisdiction is

changed, it is frequently the case that dozens of related authority records and hundreds of bib records then also have to be updated.

The CPSO page on Headings for Chinese Jurisdictions will be updated in early 2004 to reflect the changes to headings for Chinese jurisdictions that have been made in the past two years.

2. **Wade-Giles headings on bib records, identified by \$wnne and \$wnnea references**

We plan to extract from files of converted name authority records the former headings, which are coded either \$wnne or \$wnnea, and then run them against bib records in the LC database to identify headings that need to be converted.

3. **Chronological subdivisions**

Work continues on the conversion of chronological subdivisions in subject headings. A list of the subdivisions that were converted by machine appears on the pinyin home page.

4. **Systematic 041 and 043 searches on Voyager (ca. 2500 records)**

A series of searches of the 041 and 043 fields will be conducted to identify records that contain unconverted romanized Chinese strings or headings.

5. **min guo → Minguo**

When the syllables *min guo* together are used to mean the Republic of China, they must be capitalized and connected. The conversion program did not do this. There are perhaps 500 authority records and many hundreds of bib records that need to be changed.

EXAMPLE:

<u>pre-conversion WG form</u>	<u>machine converted to:</u>	<u>need change to:</u>
Chung-hua min kuo	Zhonghua min guo	Zhonghua Minguo

6. **Tibetan language bib records**

More than 1000 bib records that are coded Tibetan will be reviewed and converted. Many of the Tibetan records include romanized Chinese data in access points.

7. **"Title in Chinese"**

A search for the phrase “title in Chinese” retrieves 875 unconverted bib records in the LC database. These records will be reviewed and converted, because most of them are non-Chinese records that include romanized title added entries.

8. Unconverted access points on non-Chinese serial records

Serial records needing changes to access points will be identified in the course of performing other cleanup tasks, and sent to serials catalogers for correction.

9. Li, Po

The heading Li, Po, 701-762 converted by machine to Li, Bo, 701-762 in the autumn of 2000. The heading on authority records was changed back to Li, Po, 701-762 in March 2003. Corresponding headings on hundreds of bib records will also have to also be changed back.

10. Potential problem records identified by OCLC

LC bib records that are being flagged by the OCLC pinyin cleanup project will be sent to LC for review and conversion.

11. Other “frequently used” headings

Several other “frequently used” headings still need to be converted on non-Chinese and premarc records. They are listed at the end of this document.

CLEANUP TASKS THAT ARE NOT BEING PURSUED AT THIS TIME, BUT MAY BE EVALUATED AT A LATER DATE

1. Capitalization of generic terms for place names

The conversion program did not capitalize generic terms for place names, as called for by the romanization guidelines. This problem does not affect filing or access. These terms are now being capitalized on an as-encountered basis.

2. di / de

The conversion program automatically converted the syllable ti to di. The romanization of the character 的, therefore, converted to di rather than de. This syllable is now being changed on an as-encountered basis.

3. Bib records marked [access not affected]

Bib records that were marked for review by the conversion program have been reviewed. The many records only needing conversion of a non-access point have been marked [access not affected] in the 987 field, and have been set aside.

4. 880 fields

Portions of 880 fields sometimes did not convert, or converted differently from their parallel roman fields. Some of the reasons for this occurrence are explained in the section of the home page that describes the conversion of bibliographic records. These inconsistencies will probably be corrected on an as-encountered basis.

THE MOST FREQUENTLY USED ROMANIZED CHINESE HEADINGS

**THESE HEADINGS HAVE BEEN CONVERTED ON NON-CHINESE
AND PREMARC BIB RECORDS**

Ai, Ch'ing, 1910-
Chang, Ch'ien, 1853-1926
Ch'ang-ch'un shih ti fang chih pien tsuan wei yüan hui
Chang, Hen-shui, 1895-
Chang, T'ien-i, 1906-
Chao, Kang, 1929-
Chao, Shu-li
Ch'en, Ying-chen
Ch'en, Yün, 1905-
Cheng, Chen-to, 1898-1958
Ch'eng-tu ti t'u ch'u pan she
Chia, P'ing-wa
Chiang, Ping-chih, 1904-
Chin, Sheng-t'an. 1608-1661
Chin tai Chung-kuo shih liao ts'ung k'an
China (Republic : 1949-). Chu chi ch'u
China (Republic : 1949-). Nei cheng pu
China. Ti chih k'uang ch'an pu. Shu k'an pien chi shih
Ch'ing tai chuan chi ts'ung kan
Ching wei wen hua t'u shu ch'u pan she. Pien chi pu
Chou, En-lai, 1898-1976
Chou, Tso-jen, 1885-1967
Chou, Yang, 1908-
Chu, Hsi, 1130-1200
Ch'ü, Yüan, ca. 343-ca. 277 B.C.

Ch'ü, Pao-k'uei
 Chuang-tzu
 Chuang-tzu. Nan-hua ching
 Chung-hua ching chi yen chiu yüan ching chi chuan lun
 Chung-hua wen hua fu hsing yün tung t'ui hsing wei yüan hui
 Chung-hua wen hua ts'ung shu
 Chung-hua jen min kung ho kuo ti fang chih ts'ung shu
 Chung i ku chi cheng li ts'ung shu
 Chung-kuo fang chih ts'ung shu
 Chung-kuo fang chih ts'ung shu. Hua chung ti fang
 Chung-kuo i hsüeh pai k'o ch'üan shu
 Chung-kuo kung ch'an tang
 Chung-kuo kuo ch'ing ts'ung shu
 "Chung-kuo kuo ch'ing ts'ung shu—Pai hsien shih ching chi she hui tiao
 ch'a" pien chi wei yüan hui
 Chung-kuo ku tien wen hsüeh tso p'in hsüan tu
 Chung-kuo shao shu min tsu she hui li shih tiao ch'a tzu liao ts'ung k'an
 Chung-kuo tang tai wen hsüeh yen chiu tzu liao
 Chung-kuo ti 2 li shih tang an kuan
 Chung-kuo ti t'u ch'u pan she
 Chung-kuo wen hua shih chih shih ts'ung shu (Taipei, Taiwan)
 Chung yang yen chiu yüan
 Chung yang yen chiu yüan. Chin tai shih yen chiu so
 Fan, Wen-lan, 1891-1969
 Feng, Chi-ts'ai
 Feng, Meng-lung
 Feng, T'ien-yu, 1942-
 Feng, Yü-hsiang, 1882-1945
 Feng, Yu-lan, 1895-
 Fu, Pao-shih, 1904-1965
 Hao-jan, 1932-
 Heng-t'ang-t'ui-shih, 1711-1778
 Hsia, Yen, 1900-
 Hsiao, Hung, 1911-1942
 Hsiao-hsiao-sheng
 Hsiao-hsiao-sheng. Chin P'ing Mei tz'u hua
 Hsiao hsüeh sheng wen k'u
 Hsieh, Ling-yün, 385-433
 Hsieh, Wan-ying, 1902-
 Hsing cheng yüan wen hua chien she wei yüan hui (China)
 Hsing cheng yüan yen chiu fa chan k'ao ho wei yüan hui (China)
 Hsüan-tsang, ca. 596-664
 Hsün-tzu, 340-245 B.C.
 Hu, Shih, 1891-1962
 Huang kuan ts'ung shu
 Hui-neng, 638-713

Hung lou meng
K'ang, Yu-wei, 1858-1927
Kao, Yü-jen
“Ku pen hsiao shuo chi ch'eng” pien wei hui
Ku tai wen shih ming chu hsüan i ts'ung shu
Kuan, Chung, d. 645 B.C.
Kuan, Shan-yüeh, 1912-
Kuang-tung sheng ti t'u ch'u pan she
Kung-sun, Yang, d. 338 B.C.
Kuo hsüeh chen chi hui pien
Kuo hsüeh ming chu chen pen hui k'an
Kuo li chung yang t'u shu kuan (China)
Kuo li ku kung po wu yüan
Kuo li ku kung po wu yüan. Pien chi wei yüan hui
Kuo li pien i kuan
Kuo, Mo-jo, 1892-1978
Lao, She, 1899-1966
Lao-tzu
Lao-tzu. Tao te ching
Li, Ch'ing-chao, 1081-ca. 1141
Li, Fei-kan, 1905-
Li, Hung-chang, 1823-1901
Li shih hsiao ku shih ts'ung shu
Li, Tse-hou
Li, Yu, d. 804
Liang, Ch'i-ch'ao, 1873-1929
Lien ho pao ts'ung shu
Lin, Piao, 1908-1971
Liu, Hai-su, 1896-1994
Liu, I-sheng, fl. 1963-
Liu, Shao-ch'i, 1898-1969
Lu, Chiu-yüan, 1139-1193
Lu, Hsün, 1881-1936
Mao hsieh ts'ung k'an. Shih ch'ang yen hsi lieh
Mao, Tse-tung, 1893-1976
Nan Hua ch'u pan she
Ni, K'uang
Nieh, Jung-chen, 1899-
Ou-yang, Hsiu, 1007-1072
Ou-yang, Hsün, 557-641
Ou-yang ... (Other persons with this surname)
Pa, Chin, 1905-
Pai, Chü-i, 772-846
Pai, Hua
Pan, Ku, 32-92
Ping, Hsin, 1907-1966

Po-yang, 1920-
 San min chu i li lun ts'ung shu
 Shen, Yen-ping, 1896-
 Shu ching
 Shu, Ch'ing-ch'un, 1898-1966
 Ssu k'u ch'üan shu ... (any headings after these first 4 words)
 Ssu-ma, Ch'ien, ca. 145-ca. 86 B.C.
 Ssu-ma, Kuang, 1019-1086
 Sun-tzu, 6th cent. B.C.
 Sun-tzu, 6th cent. B.C. Sun-tzu ping fa
 T'an, Ssu-t'ung, 1865-1898
 Tang tai Chung-kuo ts'ung shu
 Tang tai Chung-kuo ts'ung shu pien chi pu
 T'ao, Hsing-chih, 1891-1946
 T'ao, Pai-ch'uan, 1903-
 Teng, Hsiao-p'ing, 1904-
 Ti t'u ch'u pan she
 T'ien, Han, 1898-1968
 Ting, Ling, 1904-
 Ts'ai, Tun-ming, 1868-1940
 Ts'ai, Yüan-p'ei
 Ts'ang hai ts'ung kan
 Ts'ao, Hsüeh-ch'in, ca. 1717-1763
 Ts'ao, Hsüeh-ch'in, ca. 1717-1763. Hung lou meng
 Ts'ao, Yü
 Ts'en, K'ai-lun
 Tseng, Kuo-fan, 1811-1872
 Tso, Tsung-t'ang, 1812-1885
 Tsu kuo ts'ung shu
 Tu, Fu, 712-770
 Tun-huang manuscripts
 Tz'u (subject hdg)
 Tzu chih t'ung chien
 Wang, Shou-jen, 1472-1529
 Wang, Yang-ming, 1472-1529
 Wen, T'ien-hsiang, 1236-1283
 Wen shih che hsüeh chi ch'eng
 Wu, Ching-tzu, 1701-1754
 Wu chiu pei chai I ching chi ch'eng
 Wu, P'ei-fu, 1874-1939
 Yeh, Sheng-t'ao, 1893-
 Yeh, Yung-lieh
 Yen, Hsi-shan, 1883-1960
 Yin te
 Yü, Fei-an
 Yü, Kuang-yüan

Yü, Yu-jen, 1878-1964

THESE HEADINGS STILL NEED TO BE CONVERTED ON NON-CHINESE AND PREMARCO BIB RECORDS

Ch'ien-lung, Emperor of China, 1711-1799

Chou li

Ch'un ch'iu

Chung yung

Han, Fei, d. 233 B.C.

Han, Fei, d. 233 B.C. Han Fei-tzu

Hsiao ching

Hsi yu chi

I ching

I li

Li chi

Mencius. Meng-tzu

Mo, Ti, fl. 400 B.C.

Mo, Ti, fl. 400 B.C. Mo-tzu

Pu-i, 1906-1967

Shan hai ching

Shih ching

Shih san ching

Shui ching chu

Shui hu chuan

Ssu shu

Ta hsüeh

Tso chuan

Yung-cheng, Emperor of China, 1677-1735